修士論文概要書

2010 年 2 月提出

専攻名 (専門分野)	情報理工学専攻	氏名	佐藤博志	指導 教員	村岡洋一　　印
研究指導名	村岡洋一	学籍番号	CD 5108B058-1		
研究 題目	未来型ライブコンサートに向けた音楽映像同期システムの開発				

1.はじめに

近年,映像メディアと音楽とは切っても切れない関係になりつつある.これは単純に,広く大衆化されたブロードキャストが映像と音楽を同時に配信していることが最大の要因である.これによって,人間の音楽に対する認識は,音楽それ自体だけではなく,それと同時に提供された映像も含めたものに変化してきている.

一般に,オーディオシステムの良し悪しを議論する場合,普通はその物理的・電気的特性を問題にするが,本当は耳に入った音をそのまま聴いているわけでなく,他の情報,過去の経験などと照らし合わせるなどの脳内処理を経た後,最終的に音として認識することが最近の研究で確認されている.つまり,ライブコンサートにおける感動をより強くするためには,音楽情報を単体として提供するだけではなく,音楽と強い関連性のある映像も同時に提供することが効果的であると考えられる.

しかし現代のライブシーンにおいては,従来から音響的な特質及び技術が重視される一方で,人間の認識を構成するもう一方の要素である映像に対する認識は高いとは言えない.そこで,観客のもつ音楽の体験,並びに映像の体験を同時に呼び起こすシステムとして,音楽と強く関係を持つ映像と,ライブで流される音楽を自動的に同期させる新規のシステムを提言する.

2.概要

本研究では,ライブコンサートを演奏と同期的な映像表現を伴った場へと進化させることを目標にしている.既存のライブコンサートでは映像表現が未だ軽視されており,単純なカメラワーク,或いは VJ(ヴィジュアルジョッキー)によるループ映像などが見られるものの,演奏と同期した形での映像表現は実現されていない.その為に,予め演奏と時間的に対応付けられた映像を,演奏速度等が異なるライブコンサートでの演奏に自動的に同期させることを目指す.

自動的な映像同期を実現する為に,まずは音楽信号の照合を行う.リハーサル音源と本番での演奏を比較する際に,演奏速度の変化や演奏誤り,加法性雑音といった照合の阻害要因を低減させた上で高速かつ高精度にアラインメントを取得する必要があり,尚且つその結果を基に映像を滑らかに出力させる必要がある.

まず初めに,リハーサルと本番での演奏との照合精度を上げる為,音楽信号の特徴抽出を高精度で行う必要がある.本研究では,主に音声認識で用いられる特徴量を基に,数多く存在する特徴量の中でどれが最も適しているかを,実際のライブシーンを想定した音源によって網羅的に検証した.

次に,高速でアラインメントを取る手法として,オンライン・タイム・ワーピング(Online Time Warping,以下 OLTW と略記)アルゴリズムに着目し,これを単にシステムに適用しただけでなく,そのアルゴリズム特有の欠点を発見し,尚且つ改良した.

最後に,アラインメント結果を基にした映像出力を異なる手法で実現し,それらに対して精度評価を行い,本システムへの適用可能性がある実装方法について提言する.

3.特徴量別精度検証実験

3.1 特徴量別精度検証の意義

音声認識の分野では以前から音響信号の特徴量を抽出し,その分析結果を元に様々な研究がなされてきた為,モデルに対して適した特徴量が経験則的に分かっている.しかし,音楽音響信号に対して特徴量を抽出する際に,最も適した手法はどのようなものであるかを特徴量別,あるいは考えられる状況別に網羅的に調べた文献が存在しない.本研究では,現行のアラインメント研究における特徴抽出方法に着目し,より高い精度を持った特徴抽出方法を調査・検討する.その際,雑音・伸縮・空白といった,ライブコンサートで発生するエラーに特に着目し,それらに対する耐性を定性的及び定量的に検証する.

3.2 提案手法

音楽情報を解析するために,楽曲の特徴量を抽出する.次に,抽出した特徴量に関して,オフラインの動的時間伸縮アルゴリズムによってアラインメントを取る.これを定性的かつ定量的に解析し,当該の特徴抽出方法の妥当性を検証する.

今回,実験に使用した特徴量は次の通りである.

・Mel Frequency Cepstral Coefficients(MFCC):12 次元
・Mel Frequency Cepstral Coefficients(MFCC):24 次元
・Linear Prediction Filter Coefficients(LPC):64 次元
・LPC Cepstral Coefficients(LPCC):12 次元
・Linear Mel-filterbank channel outputs(MELS):24 次元
・Linear Mel-filterbank channel outputs(MELS):80 次元
・Log Mel-filterbank channel outputs(FBANK):24 次元
・Log Mel-filterbank channel outputs(FBANK):80 次元

また,実験データ作成の為,曲に対して加えたエラー処理は以下の通りである.

・SN:歓声及び拍手(Small Noise)
・BN:別の曲を同音量で重ねて流した場合(Big Noise)
・IM:最初に空白(Initial Mute)
・MM:途中に空白(Middle Mute)
・CS:速度が非線形に変化(Changed Speed)

3.4 実験結果
雑音と空白に対して最も優れた特性を持つのは,MFCCとゼロ平均正規化した線形メルフィルタバンクであり,速度変化に対して最も強い耐性を持つのは MFCC であることが判明した.以上のことから,今回試した特徴量の中では,MFCC がライブコンサートには最も適用可能性が高いことが分かった.

4.OLTW アルゴリズムとその改良
4.1OLTW アルゴリズムの概要
OLTW アルゴリズムとは,動的時間伸縮処理における配列要素の更新を選択的に行うことで,実時間入力に対応させたものである.通常,動的時間伸縮アルゴリズムは比較するデータの特長量を全て費用行列に格納した上で最適経路を探索するが,一方 OLTW アルゴリズムは局所的な配列計算及び更新を連続的に行うことで最適経路の探索を行う.例えば,一辺$\mathbf{n}$の二次元配列を考えた場合に,通常の動的時間伸縮アルゴリズムは最短経路探索に必要な計算オーダーが$\mathbf{O}(n^2)$になるが,OLTW アルゴリズムを用いた場合,ユーザによって設定された一辺が$\mathbf{m}$の正方形に対して$\mathbf{O}(mn)$の計算回数で済ませることができる.これにより,低パフォーマンスな CPU を搭載したマシンでも,実時間で音楽信号のマッチングを行うことが可能になった.

4.2OLTW アルゴリズムの問題点
OLTW アルゴリズムは計算量及び実時間入力に対応した点で優れたアルゴリズムであると言えるが,その精度という意味では決して十分に高いとは言えない.例えば,正方形の一辺の長さを 40 としてみると,オフラインの場合と比較して,平均的に 199 ミリ秒(凡そ 0.2 秒)の遅延が観測された.これでは実際のライブコンサートにおいて観客が常に演奏と映像の間に 0.2 秒のズレを観測することになる.
これの原因として,OLTW アルゴリズムの暗中模索的な動き方の問題が挙げられる.オフラインの動的時間伸縮はバックトラックによって最適な経路を探しているが,OLTW アルゴリズムは二次元配列における移動経路上の「次の要素値」しか判断基準が無い.その結果として,計画性の無い動きがミスアラインメントの大きな原因となっている.

4.3 将来経路予測アルゴリズム
OLTW の誤った動きの原因となる暗中模索的な動きを抑制する為,将来経路予測アルゴリズムを実装した.これは,過去の動きを優先順位無しのキュー形式で更新して予測行列に保存し,それによって確率変数を計算することによって移動方向に補正をかけるものである.その際の確率変数の計算式は次式で与えた.

(1)将来経路予測アルゴリズムⅠ:標準型

$$P(D,V,H)=\frac{\sum_{n=1}^{elem.}\mathrm{PM(D,V,H)}}{\mathrm{elem.}}$$

ここで,各記号の表す意味は以下の通りである.
·D,V,H:方向(Diagonal,Vertical,Horizontal)
·P(方向):その方向への遷移確率
·PM(方向):予測行列内の方向要素
·elem.:予測行列の要素数

(2)将来経路予測アルゴリズムⅡ:雑音耐性強化型
誤った直角の移動を軽減させる為に,予測行列内の方向成分に対し独立的に補正をかけるのではなく,縦方向・横方向を 1 組と捉えて斜め方向への移動へと置き換える手法を実装した.

$$P(H)=\frac{\sum_{n=1}^{elem.}\mathrm{PM(H)}-\min\{\sum_{n=1}^{elem.}\mathrm{PM(V)},\sum_{n=1}^{elem.}\mathrm{PM(H)}\}}{\mathrm{elem.}}$$

$$P(V)=\frac{\sum_{n=1}^{elem.}\mathrm{PM(V)}-\min\{\sum_{n=1}^{elem.}\mathrm{PM(V)},\sum_{n=1}^{elem.}\mathrm{PM(H)}\}}{\mathrm{elem.}}$$

4.4 実験結果
予測アルゴリズムが無い OLTW は 199 ミリ秒程度の時間誤差が平均的に観測されたが,予測アルゴリズム付きの場合は平均時間誤差が約 12 ミリ秒(予測行列長:9 以上)まで減少した.2 種類の経路予測アルゴリズムの精度差は 1 ミリ秒に満たなかった為,どちらの方法で実装しても大きな変わりは無いことが分かった.また,予測行列長が 8 までは平均時間誤差が緩やかに改善する一方で(予測行列長が 8 の際は平均時間誤差が 110 ミリ秒),十分な精度向上が見られなかったが,9 を超えてからは,上述した様に約 12 ミリ秒まで時間誤差が減少し,また誤差の値は予測行列長が増えても変化しないことが確認された.

5.映像同期手法の実装と改良
映像同期手法に関しては,Java が提供する JMF では速度を引数として渡す際にレイテンシが発生し,これが蓄積することで演奏時間の半分程度の大きな遅延が生じることが確認された.このため映像を静止画に切り分けてバッファリングし,高速で出力することで動画化した.これによって時間遅延が完全に消失し,主観的にも映像同期が行えていることが確認できた.

6.まとめ
既存の資料が存在しない”ライブコンサートのアラインメント取得に適した特徴量”を,網羅的な精度検証実験によって発見した.また音楽信号照合時に,OLTW アルゴリズムによって計算オーダーを 1 次元低く抑えたまま,将来経路予測アルゴリズムをこれに追加することで,発生する時間誤差を平均 199 ミリ秒から約 12 ミリ秒へと大きく減少させ,その精度を改善することに成功した.更に,時間遅延が生じない映像同期手法を提案・実装し,滑らかな映像同期を確認した.